ECOLOGIA BALKANICA

2025, Vol. 17, Issue 2 December 2025 pp. 115-124

In silico methods for mutagenicity prediction

Vesselina Paskaleva*, Gergana Cokova

Paisii Hilendarski University of Plovdiv, Faculty of Chemistry, Department of Analytical Chemistry and Computer Chemistry, Tsar Assen Str. 24, Plovdiv, 4000, BULGARIA
*Corresponding author: vessy@uni-plovdiv.bg

Abstract. This research presents a thorough exploratory data analysis to develop an in silico model for mutagenicity prediction, contributing to Safe-by-Design strategies. Using a publicly available dataset, chemical structures were encoded via a range of molecular fingerprints and descriptors. Multiple machine learning algorithms—including k-nearest neighbors, support vector machines, and random forest—were assessed. Performance was validated through 10-fold cross-validation and further tested on an external dataset. Random Forest emerged as the most effective method, achieving a cross-validation MCC of 0.68. The in-house models showed competitive performance relative to existing publicly available tools.

Key words: ames mutagenicity, machine learning, QSAR, in silico.

Introduction

The World Health Organization warns that air pollution is an escalating danger to both ecological systems and human health. Its sources are diverse, ranging from natural phenomena such as volcanic eruptions to human-driven activities like energy production, transportation, industrial processes, agriculture, etc. Key pollutants of concern include particulate matter, carbon monoxide, ozone, nitrogen dioxide, sulphur dioxide, polycyclic aromatic hydrocarbons (PAHs), and others. PAHs comprise a broad class of organic compounds, some of which exhibit mutagenic and genotoxic properties. Their presence in the atmosphere has been linked to breast cancer, reduced lung function, obstructive respiratory diseases, etc. (Air Quality, Energy and Health, n.d.). Mutagenicity, as an ecotoxicological property of certain pollutants, adversely affects ecosystems and serves as a critical endpoint in water quality monitoring (Vargas et al., 1993). The Scientific Committee on Health and Environmental Risks of the European Commission focuses its work on evaluating the toxicity and ecotoxicity of chemicals and biochemical substances that may pose risks to human health and the environment (Scientific Committee on Health, Environmental and Emerging Risks (SCHEER) - Public Health, n.d.). In response to escalating pollution, regulatory agencies have introduced key policies like the European Green Deal (The European Green Deal, n.d.), the EC-CSS (Chemicals Strategy for Sustainability Towards a Toxic-Free Environment, n.d.), and the Zero Pollution Action Plan (European Commission. (2021), n.d.). Successfully managing the harmful effects of pollutants requires early recognition of ecological threats. The Safe and Sustainable by Design framework (Apel et al., 2024; Caldeira et al., 2022; European Commission (2022) n.d.), developed by the Joint Research Centre of the European Commission, aims to proactively prevent risks to human health and the environment. It does so by embedding safety and sustainability principles into the earliest stages of chemical development, rather than attempting to manage hazards after products have already entered ecosystems.

A valuable tool for the early identification of adverse effects is the use of in silico methods to predict chemical properties and biological acti-

vities. These computational approaches allow researchers to characterize molecules, substances, and products that have not yet been synthesized, providing insights into their potential mutagenic, toxic, or harmful effects in advance. Importantly, such methods align with the 3R (Russell & Burch, 1959; The 3Rs - The 3Rs Collaborative, n.d.) principles of animal welfare and the objectives of green chemistry.

In the present work, we describe our efforts to develop an in silico model for mutagenicity prediction. An extensive exploratory analysis was conducted, involving a wide range of molecular structure encodings and machine learning techniques. The most promising models were optimized and benchmarked against publicly available mutagenicity prediction tools. The final model was further validated using a dataset of polycyclic aromatic hydrocarbons (PAHs). Importantly, the model relies exclusively on open-source tools, ensuring alignment with the principle of reproducibility.

Materials and methods

Dataset – source and pre-processing

The molecular structures and their experimentally determined biological activities were sourced from the dataset published in Helma et al. (2021). This dataset comprises 8309 compounds, represented as SMILES linear notation. Before machine learning modeling, structural preprocessing was performed using OpenBabel. The preprocessing steps included nitro group normalization, removal of isotopic and chiral information, duplicate filtering via InChI, and retention of only the largest fragment in multi-component structures.

Preparation of the training and test sets

To ensure balanced representation, the preprocessed dataset was split into training and test sets using stratified sampling (80:20). A KNIME workflow (Fig. 1) was developed to facilitate this process. It started with a reading node to import the dataset, followed by a node that separated mutagenic and non-mutagenic molecular structures into two distinct subsets. Each subset was then randomly partitioned in an 80:20 ratio. The training set comprised 80% of both mutagenic and non-mutagenic molecules, while the test set included the remaining 20% from each category. Finally, the resulting sets were shuffled to prevent ordering bias and saved as separate files.

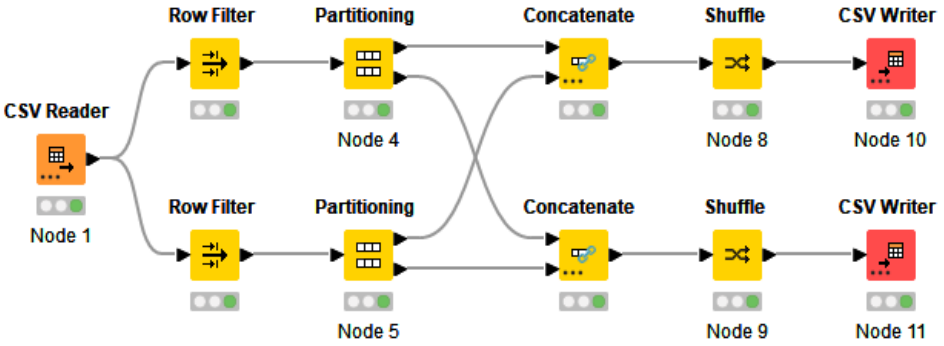

Fig. 1. KNIME workflow for stratified dataset splitting.

Molecular structure coding

Molecular structures were encoded using a variety of descriptors and fingerprints. Two software tools facilitated this process: the commercial Dragon v7.0 and the open-source PaDEL-Descriptors v2.21. Dragon was employed to calculate 0D, 1D, and 2D molecular descriptors, as well as two fingerprint types—extended circular fingerprints (ECFP) and path-based fingerprints (PFP). ECFP fingerprints were generated with bit lengths of 1024 and 2048 and a maximum radius of 2, while PFP fingerprints used the same bit lengths with a radius of 6. Atom-level features included atom type, aromaticity, total connectivity, charge, and bond order. PaDEL-Descriptors was used to compute eight fingerprint types: Fingerprinter, Extended Fingerprinter, Estate Fingerprinter, Substructure Fingerprinter, GraphOnly Fingerprinter, MACCS Fingerprinter, Pubchem Fingerprinter, and KlekotaRoth Fingerprinter.

Exploratory data analysis and model building

Exploratory data analysis (EDA) and model development were carried out using the open-source software Weka. The overall modelling workflow is illustrated in Fig. 2 and follows four main steps: (1) detailed data exploration (EDA), (2) evaluation and selection of the optimal molecular structure coding (MSC) and machine learning method (ML) pairing, (3) model refinement, and (4) final reporting.

Two MSC types were explored—Type 1, comprising fingerprints generated using an open-source tool, and Type 2, consisting of molecular descriptors and fingerprints obtained from commercial software (as previously described). Various ML algorithms were applied to both MSC types, including k-nearest neighbours, linear discriminant analysis (LDA), artificial neural networks, support vector machines (SVM), and random forest. All models were trained using default Weka settings. LDA models were built using descriptors, while all others relied on fingerprint data.

Model robustness was assessed through a 10-fold cross-validation procedure and external testing. Key statistical metrics such as Matthews's correlation coefficient (MCC), Cohen's Kappa, and others were calculated to quantify model quality. All analyses were performed on a Windows-based machine with an Intel i5-3320M processor and 8 GB of RAM.

Optimization of the selected models

To boost the performance of the selected models, we optimized the number of decision trees used in the algorithm. The initial configuration employed 100 trees. We explored a range of tree counts—specifically 20, 40, 60, 80, 100, 120, 140, 160, 180, and 200—by constructing a separate model for each value. Then we compared them based on their statistical outcomes to identify the most effective configuration.

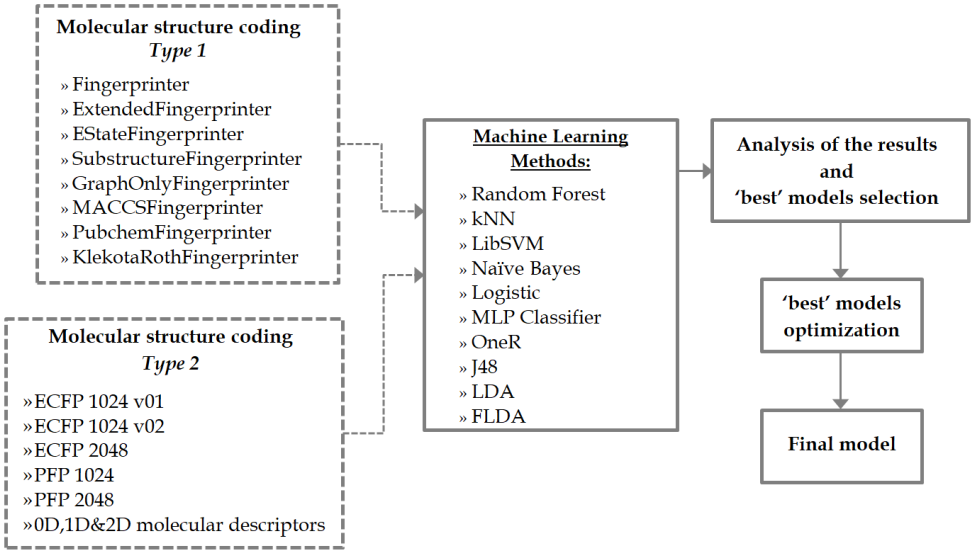

Fig. 2. Model-building workflow.

Comparative analysis

The predictive capacity of the optimized models was compared with the results obtained from two software tools — T.E.S.T., developed by the United States Environmental Protection Agency (EPA), and VEGA, developed by the Mario Negri Institute for Pharmacological Research in Italy. The T.E.S.T. software implements three models for Ames mutagenicity prediction – Nearest neighbour, Hierarchical clustering, and a Consensus model – all of which were used in the comparison. From the VEGA platform, four models (CEASAR, ISS, KNN, and SARPY) were used in the comparison and one consensus model. The comparison was made mostly on the Matthews Correlation Coefficient (MCC).

Used software

Molecular structure preprocessing was performed using OpenBabel v3.1.1 (Open Babel - the

Chemistry Toolbox — Open Babel Openbabel-3-1-1 Documentation, n.d.). Dataset partitioning into training and test sets was carried out via a custom-built workflow in KNIME v4.5.1 (KNIME AG. (2021). KNIME Analytics Platform (Version 4.5.1) [Computer Software]. KNIME AG, n.d.). Fingerprints and descriptors were computed using the open-source PaDEL Descriptors v2.17 (Yap, 2011). Additionally, 0D, 1D, and 2D molecular descriptors, along with Extended Connectivity and Path fingerprints, were calculated using the commercial software Dragon v7.0. Exploratory data analysis and model development were conducted using Weka v3.8.6 (University of Waikato, 2022). Weka: The Waikato Environment for Knowledge Analysis (Version 3.8.6) [Computer Software]. University of Waikato, n.d.). The resulting models were compared against existing mutagenicity prediction tools, including VEGA v1.1.5-b48 (Benfenati et al., n.d.) and T.E.S.T. v5.1.2.0 (Toxicity Estimation Software Tool (TEST) | US EPA, n.d.).

Results

After the pre-processing, the dataset consists of 8006 chemical structures - 3983 mutagenic and 4023 non-mutagenic. Their structural characteristics are visualized in terms of molecular weight (Fig. 3A), heavy atom count (Fig. 3B), and ring system count (Fig. 3C).

Using stratified splitting, the pre-processed dataset was divided into a training set of 6403 molecules (3186 mutagenic) and a test set of 1602 molecules (797 mutagenic). Molecular weight distributions for both sets are shown in Fig. 4.

Molecular descriptors and fingerprints were generated for both the training and test datasets using Dragon v7.0 and PaDEL-Descriptors. Dragon v7.0 produced a total of 1835 descriptors spanning 0D, 1D, and 2D levels, along with two types of fingerprints—Extended Connectivity Fingerprints and Path-based Fingerprints—each computed at lengths of 1024 and 2048 bits. PaDEL-Descriptors contributed eight fingerprint formats: Fingerprinter (1024 bits), Extended Fingerprinter (1024 bits), Estate Fingerprinter (79 bits), Substructure Fingerprinter (307 bits), GraphOnly Fingerprinter (1024 bits), MACCS Fingerprinter (66 bits), Pubchem Fingerprinter (881 bits), and KlekotaRoth Fingerprinter (4860 bits).

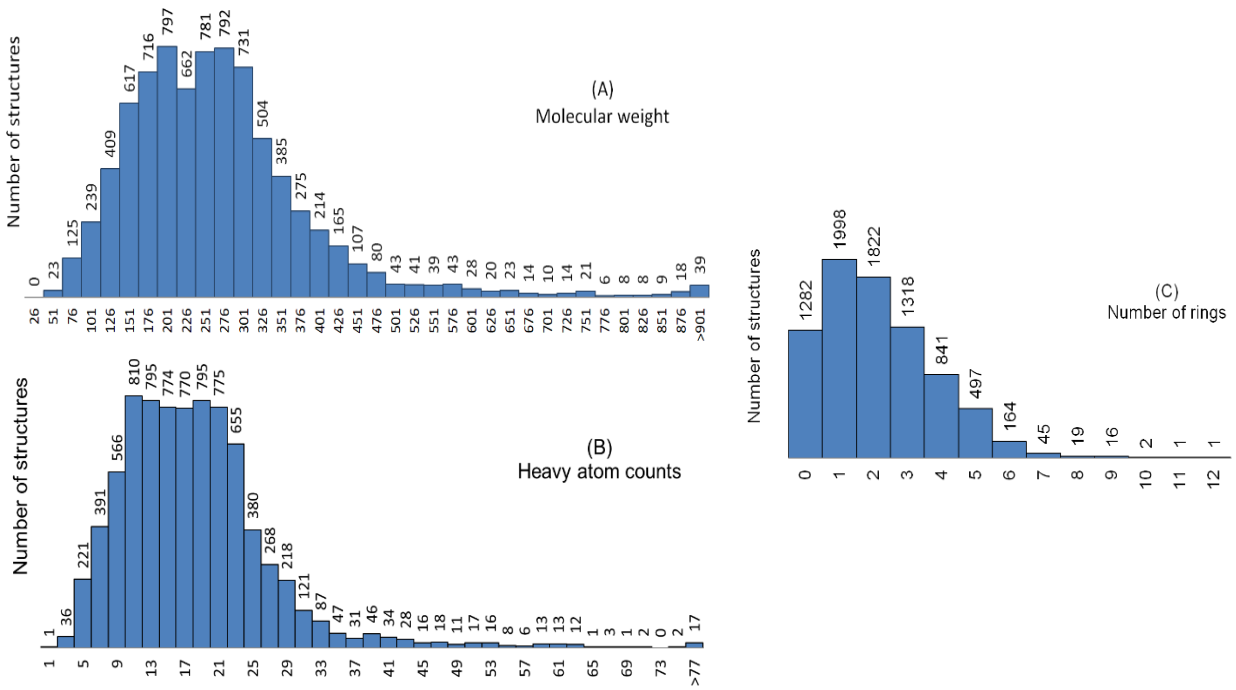

Fig. 3. Distribution of the structures with respect to their MW (A), Heavy atom count (B), and number of rings (C).

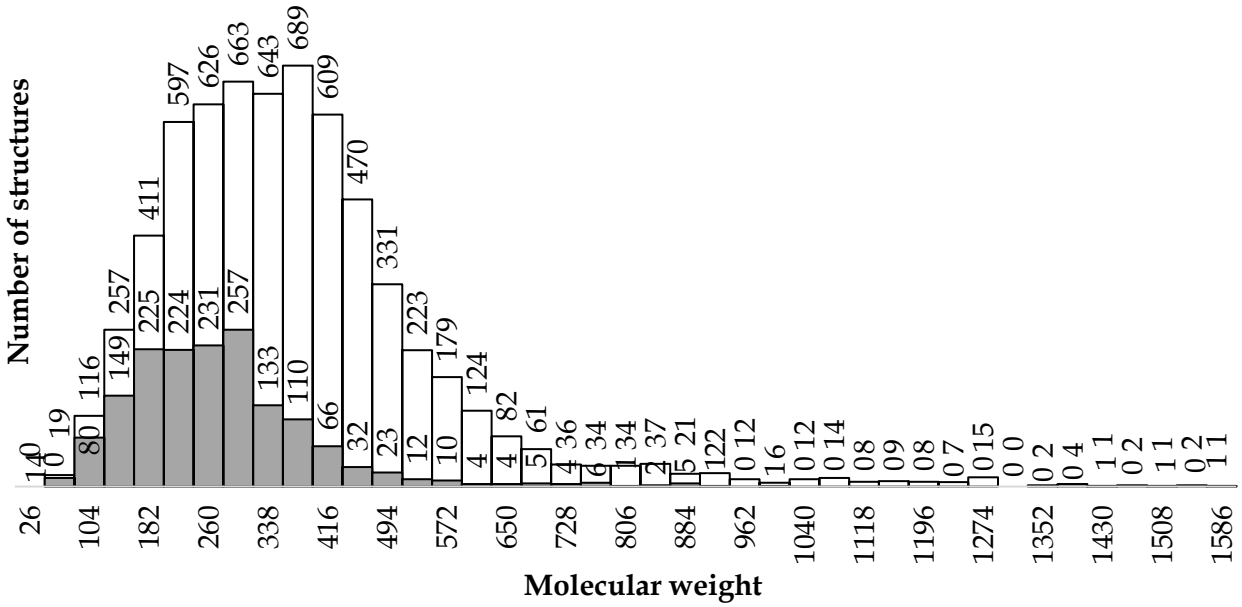

Fig. 4. Distribution of the molecules from the train (in white) and test (in gray) sets according to their molecular weight.

As part of the exploratory data analysis, we developed over 150 models by combining different molecular structure encodings with various machine learning algorithms. Each model was validated using 10-fold cross-validation and then tested on an external dataset. The performance, measured by MCC values across training, validation, and test sets, is detailed in Table 1.

Table 1. Range of MCC values obtained for each ML set for training, validation, and testing.

ML method	Number of models	Training	CV10	Test set
Random Forest	21	[0.810; 0.996]	[0.536; 0.683]	[0.558; 0.707]
iBk	21	[0.811; 0.995]	[0,515; 0,622]	[0.563; 0.650]
LivSVM	21	[0.361; 0.992]	[0.298; 0.564]	[0.296; 0.572]
Naïve Bayes	20	[0.235; 0.409]	[0.221; 0.389]	[0.227; 0.417]
Logistic	5	[0.744; 0.995]	[0.493; 0.556]	[0.474; 0.793]
MLP Classifier	3	[0.819; 0.938]	[0.556; 0.567]	[0.524; 0.581]
OneR	21	[0.241; 0.494]	[0.237; 0.286]	[0.216; 0.278]
J48	20	[0.641; 0.938]	[0.495; 0.611]	[0.514; 0.632]
LDA	14	[0.427; 0.836]	[0.416; 0.585]	[0.403; 0.605]
FLDA	14	[0.429; 0.836]	[0.418; 0.586]	[0.404; 0.605]

To identify the strongest models, we ranked all results by MCC in descending order, ensuring the best-performing ones were listed first. This ranking method was repeated for every statistical parameter calculated. A summary of the top five models for each metric is provided in Table 2.

Selected models

Two models were identified as the top performers and selected for further optimization. Model 1 is based on molecular structure coding (MSC) using ECFP_1024_v02, while Model 2 utilizes Pubchem fingerprints. Both models employ the Random Forest algorithm. For Model 1, the Matthews Correlation Coefficient values were 0.993 for training, 0.683 for validation, and 0.704 for testing. Model 2 achieved MCC values of 0.979, 0.662, and 0.676 for the respective datasets.

Table 2. Ranking of the build models based on a given statistical metric.

	MCC	FNR	ROC Area	TP Rate	Kappa	CC
MCC						
ECFP_1024_v02+RF	**0.683**	0.177	0.909	0.841	0.6823	84.1193
ECFP_2048+RF	**0.678**	0.182	0.909	0.839	0.6779	83.9007
ECFP_1024_v01+RF	**0.678**	0.183	0.908	0.839	0.6776	83.8826
PubchemFP+RF	**0.662**	0.181	0.9	0.831	0.6619	83.099
PFP_2048+RF	**0.658**	0.201	0.897	0.829	0.657	82.8545
FNR						
PFP_1024+MLP Classifier	0.567	**0.171**	0.837	0.782	0.564	78.201
Descriptors_D+Naïve Bayes	0.257	**0.174**	0.688	0.616	0.233	61.581
ECFP_1024_v02+RF	0.683	**0.177**	0.909	0.841	0.682	84.119
Pubchem Fingerprint+RF	0.662	**0.181**	0.900	0.831	0.662	83.099
ECFP_2048+RF	0.678	**0.182**	0.909	0.839	0.678	83.901
ROC Area						
ECFP_1024_v02+RF	0.683	0.177	**0.909**	0.841	0.6823	84.1193
ECFP_2048+RF	0.678	0.182	**0.909**	0.839	0.6779	83.9007
ECFP_1024_v01+RF	0.678	0.183	**0.908**	0.839	0.6776	83.8826
FP-all+RF	0.656	0.186	**0.902**	0.828	0.656	82.8022
PubchemFP+RF	0.662	0.181	**0.900**	0.831	0.6619	83.099
TP Rate						
ECFP_1024_v02+RF	0.683	0.177	0.909	**0.841**	0.6823	84.1193
ECFP_2048+RF	0.678	0.182	0.909	**0.839**	0.6779	83.9007
ECFP_1024_v01+RF	0.678	0.183	0.908	**0.839**	0.6776	83.8826
PubchemFP+RF	0.662	0.181	0.9	**0.831**	0.6619	83.099
PFP_2048+RF	0.658	0.201	0.897	**0.829**	0.657	82.8545
Kappa						
ECFP_1024_v02+RF	0.683	0.177	0.909	0.841	**0.6823**	84.1193
ECFP_2048+RF	0.678	0.182	0.909	0.839	**0.6779**	83.9007
ECFP_1024_v01+RF	0.678	0.183	0.908	0.839	**0.6776**	83.8826
PubchemFP+RF	0.662	0.181	0.9	0.831	**0.6619**	83.099
PFP_2048+RF	0.658	0.201	0.897	0.829	**0.657**	82.8545
KK						
ECFP_1024_v02+RF	0.683	0.177	0.909	0.841	0.6823	**84.1193**
ECFP_2048+RF	0.678	0.182	0.909	0.839	0.6779	**83.9007**
ECFP_1024_v01+RF	0.678	0.183	0.908	0.839	0.6776	**83.8826**
PubchemFP+RF	0.662	0.181	0.9	0.831	0.6619	**83.099**
PFP_2048+RF	0.658	0.201	0.897	0.829	0.657	**82.8545**

Optimization of the selected models

Following optimization based on the number of trees, the final parameters for the selected models were established. Model 1 utilizes molecular structure coding via ECFP with a size of 1024 and employs a Random Forest algorithm configured with 140 trees. Model 2 is based on Pubchem fingerprints and uses Random Forest with 100 trees.

Comparison

The performance of the selected models was benchmarked against two established platforms that include mutagenicity prediction —VEGA and T.E.S.T.—using results obtained from an external test set. All models from VEGA and T.E.S.T., along with our developed Model 1 and Model 2, were ranked based on calculated statistical metrics. Rankings for Matthews Correlation Coefficient (MCC) and False Negative Rate (FNR) are presented in Table 3.

Table 3. Comparison vs publicly available software according to MCC (left) and FNR (right).

Model	MCC	Model	FNR
VEGA Consensus v.1.0.3	0.93	VEGA Consensus v.1.0.3	0.01
VEGA KNN v.1.0.0	0.89	VEGA KNN v.1.0.0	0.03
T.E.S.T. Hierarchical Clustering	0.73	VEGA SarPy/IRFMN v.1.0.7	0.06
Model 1	***0.70***	VEGA CEASAR v.2.1.9	0.08
VEGA CEASAR v.2.1.9	0.68	***Model 1***	***0.12***
Model 2	***0.68***	T.E.S.T. Hierarchical Clustering	0.13
T.E.S.T. Consensus	0.67	VEGA ISS v.1.0.2	0.14
VEGA SarPy/IRFMN v.1.0.7	0.60	***Model 2***	***0.15***
VEGA ISS v.1.0.2	0.55	T.E.S.T. Consensus	0.17
T.E.S.T. KNN	0.55	T.E.S.T. KNN	0.22

Application of model 'PubchemFingerprint + RandomForest (trees=100)' on polycyclic aromatic hydrocarbons

Model 2 was applied to a dataset comprising 70 polycyclic aromatic hydrocarbons (PAHs), sourced from Papa et al. (2008) and consisting of 44 mutagenic and 26 non-mutagenic compounds. The model's performance was evaluated using the resulting confusion matrix (Table 4), which served as the basis for calculating statistical parameters.

The statistical metrics resulting from the application of Model 2 to the PAHs dataset are summarized in Table 5.

Table 4. Confusion Matrix Output: *Model 2* Applied to PAHs dataset.

		Experimentally observed class	
		Mutagenic	Non-mutagenic
Predicted class	Mutagenic	TP = 40	FP = 20
	Non-mutagenic	FN = 4	TN = 6

Table 5. *Model 2* Performance Statistics on PAHs Dataset.

Parameter	Value	Parameter	Value
Accuracy	0.657	False negative rate	0.136
Balanced accuracy	0.570	Positive predictive value	0.667
MCC	0.193	Negative predictive value	0.600
Sensitivity	0.909	False discovery rate	0.333
Specificity	0.231	F1	0.769
False positive rate	0.769		

Discussion

The pre-processed dataset demonstrates significant structural diversity (Fig.3). Most compounds (70%) have molecular weights between 151 and 326, while a small fraction (~5%) are below 100, and only 0.5% exceed 900. Heavy atom counts range from 1 to 214, with nearly three-quarters of the molecules falling between 9 and 23 atoms. Very few structures (~0.3%) contain more than 70 heavy atoms. In terms of ring systems, 16% are acyclic, 25% have a single ring, 23% contain two rings, and only 0.05% feature more than ten rings. Following stratified splitting, the molecular weight distributions in the training and test sets remained comparable (Fig.4).

Among the various models developed, Random Forest consistently outperformed other machine learning techniques. Based on the Matthews Correlation Coefficient from cross-validation, the top three models were all RF-based and utilized molecular structure coding derived from ECFP fingerprints, with variations in fingerprint para-

meters. The fourth-ranked model employed PubChem fingerprints generated using an open-source descriptor calculator.

In ML applications for mutagenicity prediction, accurately identifying mutagenic compounds is crucial. Therefore, a high True Positive Rate (TPR) is desirable. However, the False Negative Rate (FNR) is equally important, as it determines how often harmful compounds are mistakenly classified as safe. Compounds predicted as mutagenic typically undergo further validation, whereas those predicted as non-mutagenic often do not—introducing the risk of overlooking harmful substances. Hence, models with low FNR are preferred.

Our top five models ranked by the Matthews Correlation Coefficient, showed strong alignment with those ranked by True Positive Rate (Table 2). False Negative Rate rankings were evaluated in ascending order to emphasize models with the lowest error in negative predictions. The model ranked highest by MCC and TPR placed third in the FNR ranking, while the fourth-ranked MCC/TPR model also ranked fourth by FNR. Notably, the top two models by FNR did not appear in the top five for any other metric.

Following this analysis, two models were selected for further optimization. The first is the top-ranked model based on MCC and TPR, employing the Random Forest algorithm with molecular structure coding derived from ECFP fingerprints, calculated using the commercial software Dragon v7.0. The second is the fourth-ranked MCC/TPR model, which also utilizes RF and MSC, but is based on PubChem fingerprints generated via the open-source tool Padel-Descriptors.

Random Forest models typically achieve higher predictive accuracy with an increased number of trees, though this comes at the cost of longer computation time. To balance predictive power and efficiency, we aimed to select a tree count that ensures strong statistical results while maintaining reasonable increases in computation time. Comparisons were based on cross-validation outcomes. Model 2 demonstrated high stability across statistical parameters. Only the ROC Area and false negative rate showed slight dips below 80 trees. The default value of 100 trees was selected as optimal due to its top ranking across most metrics. Model 1 remained statistically stable overall, but performance dropped when using fewer than 60 trees. For the optimal number of trees, we chose 140 - it ranked first in ROC Area and remained among the top four in all other metrics and without significantly increasing the number of trees.

As shown in Table 3, our model's performance is comparable to, and in some cases exceeds, that of the mutagenicity models developed by the United States Environmental Protection Agency (T.E.S.T.) and the Mario Negri Institute for Pharmacological Research (VEGA). Based on the Matthews Correlation Coefficient, both of our models outperformed T.E.S.T. Consensus, VEGA SarPy/IRFMN, VEGA ISS, and T.E.S.T. KNN. Among the ten models evaluated (five models from VEGA, three models from T.E.S.T., and two in-house), Model 1 ranked 4th and Model 2 ranked 6th overall. In terms of false negative rate, both models surpassed the performance of T.E.S.T. Consensus and KNN.

Although Model 1 outperformed Model 2 based on statistical metrics and ranked higher compared to other software, Model 2 offers a key advantage: it is entirely built on open-source tools, enhancing reproducibility. For this reason, we selected Model 2 to evaluate its applicability to a dataset of polycyclic aromatic hydrocarbons. The model demonstrated strong performance in identifying mutagenic compounds, correctly predicting 91% of them. However, its ability to detect non-mutagenic compounds was limited, with only 23% correctly classified. Notably, the model exhibited a low false negative rate (13.6%), though it produced a high number of false positives (76.9%).

Conclusions

Following an extensive exploratory data analysis involving diverse molecular structure encodings and machine learning algorithms, the Random Forest method emerged as the most effective. Models utilizing commercial fingerprinting tools outperformed those based on open-source encodings. While Model 1 slightly outperformed Model 2 in predictive accuracy, the latter offers a significant advantage in terms of reproducibility—an essential feature for in-silico applications. Both models demonstrated competitive performance when benchmarked against tools from the EPA and the Institute for Pharmacological Re-

search in Italy. Moreover, Model 2's application to an ecological dataset revealed strong mutagenicity prediction capabilities, though with room for improvement in identifying non-mutagenic compounds, highlighting targets for future refinement, supporting mutagenicity prediction within the Safe and Sustainable-by-Design framework.

Acknowledgements

This study is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project No. BG-RRP-2.004-0001-C01.

References

Air quality, energy and health. (n.d.). Retrieved October 21, 2025, from https://www.who.int/teams/environment-climate-change-and-health/air-quality-energy-and-health/health-impacts

Apel, C., Kümmerer, K., Sudheshwar, A., Nowack, B., Som, C., Colin, C., Walter, L., Breukelaar, J., Meeus, M., Ildefonso, B., Petrovykh, D., Elyahmadi, C., Huttunen-Saarivirta, E., Dierckx, A., Devic, A.C., Valsami-Jones, E., Brennan, M., Rocca, C., Scheper, J., Strömberg, E., & Soeteman-Hernández, L.G. (2024). Safe-and-sustainable-by-design: State of the art approaches and lessons learned from value chain perspectives. *Current Opinion in Green and Sustainable Chemistry*, 45, 100876. doi: 10.1016/J.COGSC.2023.100876

Benfenati, E., Manganaro, A., & Gini, G. (n.d.). VEGA-QSAR: AI inside a platform for predictive toxicology. Retrieved October 21, 2025, from http://journal.chemistrycentral.com/supplements/4/S1

Caldeira, C., Farcal, L.R., Garmendia Aguirre, I., Mancini, L., Tosches, D., Amelio, A., Rasmussen, K., Rauscher, H., Riego Sintes, J., & Sala, S. (2022). *Safe and sustainable by design chemicals and materials - Framework for the definition of criteria and evaluation procedure for chemicals and materials*. Publications Office of the European Union, JRC128591. doi: 10.2760/487955

Chemicals Strategy for Sustainability Towards a toxic-free environment. (n.d.). Retrieved October 21, 2025, from https://ec.europa.eu/environment/chemicals/index_en.htm

European Commission. (2021). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Pathway to a healthy planet for all EU action plan: Towards zero pollution. Document 52021DC0400. Available at: https://eur-lex.europa.eu/

European Commission. (2022). Commission recommendation establishing a European assessment framework for safe and sustainable by design chemicals and materials (C(2022) 8854 final). Document 32022H2510. Available at: https://eur-lex.europa.eu/

Helma, C., Schöning, V., Drewe, J., & Boss, P. (2021). A Comparison of Nine Machine Learning Mutagenicity Models and Their Application for Predicting Pyrrolizidine Alkaloids. *Frontiers in Pharmacology*, 12, 708050. doi: 10.3389/FPHAR.2021.708050/BIBTEX

KNIME AG. (2021). KNIME Analytics Platform (Version 4.5.1) [Computer software]. KNIME AG. (n.d.). Retrieved October 21, 2025, from https://www.knime.com/

Kode Chemoinformatics & Talete srl. (2017). Dragon (Version 7.0) [Computer software]. Kode Chemoinformatics. (n.d.). Retrieved October 21, 2025, from https://www.talete.mi.it/products/dragon_description.htm

Open Babel - the chemistry toolbox — Open Babel openbabel-3-1-1 documentation. (n.d.). Retrieved October 21, 2025, from https://openbabel.org/

Papa, E., Pilutti, P., & Gramatica, P. (2008). Prediction of PAH mutagenicity in human cells by QSAR classification. *SAR and QSAR in Environmental Research*, 19(1–2), 115–127. 10.1080/10629360701843482;WGROUP:STRING:PUBLICATION

Russell, W.M., & Burch, R.L. (1959). *The principles of humane experimental technique*. London: Methuen & Co., 252 p.

Scientific Committee on Health, Environmental and Emerging Risks (SCHEER) - Public Health. (n.d.). Retrieved October 21, 2025, from https://health.ec.europa.eu/scientific-committees/scientific-committee-health-environmental-and-emerging-risks-scheer_en

The 3Rs - The 3Rs Collaborative. (n.d.). Retrieved October 21, 2025, from https://3rc.org/the-3rs/

The European Green Deal — European Environment Agency. (n.d.). Retrieved October 21, 2025, from https://www.eea.europa.eu/policy-documents/com-2019-640-final

Toxicity Estimation Software Tool (TEST) | US EPA. (n.d.). Retrieved October 21, 2025, from https://www.epa.gov/comptox-tools/toxicity-estimation-software-tool-test

University of Waikato. (2022). Weka: The Waikato Environment for Knowledge Analysis (Version 3.8.6) [Computer software]. University of Waikato. (n.d.). Available at: https://ml.cms.waikato.ac.nz/weka/

Vargas, V.M.F., Motta, V.E.P., & Henriques, J.A.P. (1993). Mutagenic activity detected by the Ames test in river water under the influence of petrochemical industries. *Mutation Research / Genetic Toxicology*, 319(1), 31–45. doi: 10.1016/0165-1218(93)90028-C

Yap, C.W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*, 32(7), 1466–1474. doi: https://doi.org/10.1002/jcc.21707

Received: 23.10.2025
Accepted: 04.12.2025